

1. Use the following table about people who are in doctoral programs:

	Men	Women	
Completed	423	98	521
Still Enrolled	134	33	167
Dropped Out	238	98	336
	795	229	1024

a. Add in the totals rows/columns

b. How many rows? 3

c. How many columns? 2

d. How many variables? What are they?

2 doctoral program completion & gender

e. How many cells? 6

f. What percent of people surveyed are men? $P(M) = \frac{795}{1024} = 77.64\%$

g. What percent of people surveyed are women? $P(W) = \frac{229}{1024} = 22.36\%$

h. What percent of people surveyed completed the program? $P(C) = \frac{521}{1024} = 50.88\%$

i. What percent of people surveyed are still enrolled in the program? $P(E) = \frac{167}{1024} = 16.31\%$

j. What percent of people surveyed are dropped out of the program? $P(D) = \frac{336}{1024} = 32.81\%$

k. What percent of people surveyed are men and completed the program?

$$P(M \cap C) = \frac{423}{1024} = 41.31\%$$

l. What percent of people surveyed are women and dropped out of the program?

$$P(W \cap D) = \frac{98}{1024} = 9.57\%$$

m. What percent of those that dropped out were men? How about women?

$$P(M|D) = \frac{238}{336} = 70.83\% \quad | \quad P(W|D) = \frac{98}{336} = 29.17\%$$

n. What percent of those that completed the program are men? How about women?

$$P(M|C) = \frac{423}{521} = 81.19\% \quad | \quad P(W|C) = \frac{98}{521} = 18.81\%$$

o. What percent of men are still enrolled? How about of women?

$$P(E|M) = \frac{134}{795} = 16.86\% \quad | \quad P(E|W) = \frac{33}{229} = 14.41\%$$

2. For the table in problem 1, find the expected cell counts BY HAND (show work):

	Men	Women
Comp.	$\frac{795 \times 521}{1024} = 404.49$	$\frac{521 \times 229}{1024} = 116.51$
Still	$\frac{795 \times 167}{1024} = 129.65$	$\frac{167 \times 229}{1024} = 37.35$
Dropped	$\frac{795 \times 336}{1024} = 260.86$	$\frac{336 \times 229}{1024} = 75.14$

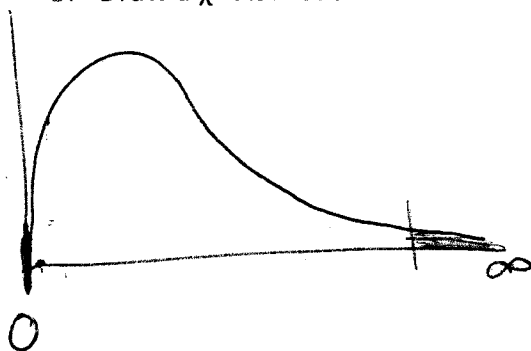
3. A large value of χ^2 tells us what about the expected and observed counts? What does a small value tell us?

that they are far apart +
that we should reject H_0
that they are close, + we will

4. What other distribution is the χ^2 distribution related to?

Z distribution - squared!

5. Draw a χ^2 distribution and describe the shape.



- right skewed

6. Is a χ^2 value of -12.3 possible? Why or why not?

no. χ^2 values are all positive numbers - they are squared values

7. Why does the χ^2 test always use the upper part (or direction) of the χ^2 distribution?

- because a p-value looks in the "more extreme" direction, and for χ^2 , more extreme is more positive.
Can't be negative

Show all appropriate work on the following problems. All tests of significance must have the following to be considered complete:

- Conditions checked
- Hypotheses (written out)
- Test statistic formula and value
- P-value (probability notation and value)
- Conclusion (full 2 sentences)

χ^2 G.O.F.

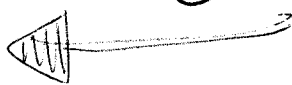
8. "Looking Up to Athletes" (USA Today, May 7, 1991) reported "Here's how sports team members say athletes do as role models for children: Excellent - 16%, Good - 38%, Fair - 41%, Poor - 5%." Suppose you took a poll of 350 members within your community and obtained the following results (in the same order): 44, 145, 133, 28. Do your results show that your community has a significantly different idea about athletes as role models than the sports team members?

<u>observed</u>	<u>expected</u>
44	56
145	133
133	143.5
28	17.5

Conditions

① SRS ✓ (assumed)

② all expected counts ≥ 5 ✓



H_0 : the observed sample distribution of community views of sports team members fits the exp. distr.

H_a : the observed sample distr. of community views of sports team members ~~fits~~ the exp. distr. doesn't fit

$$\chi^2 = \sum \frac{(\text{obs} - \text{exp})^2}{\text{exp}} = \frac{10.72}{10.72} \quad df = 3$$

$$P(\chi^2 > \frac{10.72}{10.72}) = 0.0133 \quad \alpha = 0.05$$

- We reject H_0 b/c p-value $< \alpha = 0.05$

- We have sufficient evidence that the observed sample distr. of community views of sports^{team} members doesn't fit the expected distr.

χ^2 test for association

9. Alcohol and nicotine consumption during pregnancy may harm the fetus permanently. Because drinking and smoking behaviors may be related, it is important to understand the nature of this relationship when assessing the possible effects on children. One study classified 452 mothers according to their alcohol intake prior to pregnancy recognition and their nicotine intake during pregnancy. Carry out a complete analysis of the association between alcohol and nicotine consumption.

		Nicotine (mg/day)		
		None	15-Jan	16 or more
Alcohol (oz/day)	None	105	7	11
	0.01-0.10	58	5	13
	0.11-0.99	84	37	42
	1.0 or more	57	16	17

Conditions

- 1) SRS ✓ (assumed)
- 2) all exp. counts ≥ 5 ✓
(looked in matrix B)

H_0 : alcohol & nicotine consumption are independent.

H_a : alcohol & nicotine consumption are dependent.

$$\chi^2 = \sum \frac{(\text{obs} - \text{exp})^2}{\text{exp}} = 42.252$$

$$P(\chi^2 > 42.252) = 1.64 \times 10^{-7} \quad df = 6$$

- we reject H_0 b/c p-value $< \alpha = 0.05$.

- we have sufficient evidence that alcohol & nicotine consumption are dependent

χ^2 GOF

10. A city expressway utilizing four lanes in each direction was studied to see whether drivers preferred to drive on the inside lanes. A total of 1000 automobiles were observed during the heavy early-morning traffic, and their respective lanes were recorded. Do the data present sufficient evidence to indicate that some lanes are preferred, or that they are EQUALLY preferred? State your hypotheses and carry out the test.

Lane	1	2	3	4
L_1 Observed Count	294	276	238	192
L_2 Expected	250	250	250	250

Conditions

- 1) SKS ✓
- 2) all expected counts
25 ✓
(all 250)

H_0 : the observed sample distr. of lanes fits the expected distr.

H_a : the obs. sample distr. of lanes doesn't fit the exp. distr.

$$\chi^2 = \sum \frac{(\text{obs} - \text{exp})^2}{\text{exp}} = 24.48$$

$$P(\chi^2 > 24.48) = 1.983 \times 10^{-5} \quad df = 3$$

- We reject H_0 b/c $p\text{-value} < \alpha = 0.05$
- We have sufficient evidence that...
(recopy H_a)

So... some lanes are preferred.

11. A psychologist is investigating how a person reacts to a certain situation. He feels the reaction may be influenced by how ethnically pure the person's neighborhood is. He collects data on 500 people. Does there appear to be a relationship between neighborhood and reaction at the 0.10 level of significance?

Pure?	Mild	Medium	Strong
Yes	170	100	30
No	70	100	30

Conditions

- 1) SRS ✓
- 2) all expected counts ≥ 5 ✓
(check matrix B)

H_0 : purity and reaction are independent

H_a : purity & reaction are dependent.

$$\chi^2 = \sum \frac{(\text{obs} - \text{exp})^2}{\text{exp}} = 22.5694$$

$$P(\chi^2 > 22.5694) = 1.256 \times 10^{-5} \quad df=2$$

- We reject H_0 b/c $p\text{-value} < \alpha = 0.10$
- We have sufficient evidence that...
(recopy H_a).